Document Authentication Workshop

U.S. Government Printing Office

June 18, 2010

Agenda

- Introductions and Welcome
- Authentication Background
 - Authentication for Individuals
 - Authentication for Automated, High Volume (Bulk data) applications/uses
- Standards and Methods for Bulk Data Authentication
- Chain of Custody Use Cases
- Re-authentication over Time
- Granular Authentication

Introduction

Focus for today's dialogue:

 Authentication of publicly disseminated GPO electronic documents on the Internet

What is OUTSIDE the scope of today's dialogue:

• Authentication as applied to GPO products that are not electronic documents publicly available on the Internet

Introduction

Desired Outcomes:

- Feedback from the constituencies represented here to GPO to inform GPO decisions on future system deployments
 - Input for RFI for future Industry Day with vendors at GPO
- GPO decision topics:
 - Does the community require different levels of authentication assurance on the same content (some parties willing to use less robust means and others requiring the current digital signature approach)?
 - What standards and techniques to use for native XML authentication?
 - What techniques and standards should be used for "Chain of Custody" (if content originators support it)?
 - What are the community requirements for "Granular Authentication"?

Authentication Background

- GPO publishes PDF files digitally signed using cryptographic digital signatures and PKI
 - Signature method uses open standard (IETF 2315/5652, aka PKCS#7)
 - PDF based on an open standard
- Positive feedback from all stakeholder and user communities
- Easy to use and reliable for users and citizens

Authentication Background

- GPO publishes digitally signed PDF files
 - Starting with President's Budget in FY2009 on GPOAccess web site
 - Congressional Bills starting with the 110th
 Congress on GPOAccess web site
 - FDsys Beta starting in FY2009
 - FDsys is signing all collections; as collections are published signed PDF files available

GPO Authentication Goals

- Provide assurance to document recipients or readers of:
 - Source identity (publisher) for the document
 - GPO as the source for GPO disseminated documents
- Provide assurance document not altered since publication
 Provide means of reliably detecting if document was altered
- Provide a method that supports authentic chain of custody
 - Chain of custody can be reliably provided and not altered
 - Provide a means of reliably detecting if the chain of custody was altered

Authentication Background

- Factors that seem important to future direction:
 - Authentication technique that is strong enough to maximize the length of time it is valid
 - Authentication technique that works even when disconnected from the Internet
 - Authentication technique that is based on established international, open standards (as opposed to proprietary methods)
 - Authentication technique based on binary data (rather than proprietary formats)

Authentication Background

- Factors that seem important to future direction:
 - Authentication technique that is clear and simple about the publisher of the document
 - Authentication technique that is easily extensible, using open international standards, to clearly and simply displaying for the consumer the chain of custody (provenance of the document)
- FEEDBACK ON THESE FACTORS. ARE THERE OTHER IMPORTANT FACTORS?
 - Participant feedback

Authentication Techniques Beyond PDF

- Since Cryptographic Digital Signatures used and effective for PDF → Cryptographic Digital Signatures seem natural for other data types
- Question: Is a 2nd assurance level technique like Hash based schemes (that may not do all that digital signature does) needed for a segment of the user community?
 - See Comparative Table on Next Slide
- Follow on Question: If so, what standards should guide that 2nd technique?
- FEEDBACK IS REQUESTED. IS SOME OTHER STANDARDIZED TECHNIQUE "BETTER"?
 - Participant feedback

Comparison Among Authentication Techniques

	COMPARATIVE FACTORS							
	Trusted Third Party Involved?	Could Trusted Third Party be GPO?	Does it use Open Standards?	Does the method show the publisher of the document?	Does the method facilitate "Chain of Custody"?	Free Client software available for users to validate?	Is the method included in the NIST guidance for Electronic Authentication? (NIST Special Publication 800-63)	Is offline validation possible?
Method								
Hash Based Method	Yes	Yes	Yes (for Hash)	No	No*	Yes	No	No
Cryptographic digital signature	Yes (Root CA)	Yes	Yes (IETF RFC 5652 for binary data; W3C standards for XML data)	Yes	Yes	Yes	Yes	Yes

*=No standards exist for this function

Authentication Techniques Beyond PDF

- An option is to use PDF as a "carrier" to encapsulate other file types
 - Embed (encapsulate) other file types inside of a PDF file
 - Disadvantages:
 - Embedded files don't have direct authentication it would be tied to the PDF file
 - No standards for linking the authentication information between all the files
- GPO conclusions, to this point:
 - Not the best option
 - Examine other options (next slides)

Authentication Standards for Automated, High Volume applications

Options :

- World Wide Web Consortium (W3C) XML authentication standards
 - Based on cryptographic digital signature
- IETF 5652 (PCKS #7)
 - Uses cryptographic digital signature
 - Is used by many other standards (S/MIME v3, IETF RFC 3851, for example)
- FEEDBACK FROM PARTICIPANTS
 - Input from participants

Options :

- Enveloped Signature
 - Signature is embedded within the document containing the signed content
- Enveloping Signature
 - Signature contains the signed content itself, all within the document
- Detached Signature
 - Signature is separate from the content
- Pros and Cons of these approaches on the following slides

Enveloped Signature

Signature is embedded within the document containing the signed content

Consequences:

- The signature is a Child of the content
- Document needs a placeholder to hold the signature

Advantages:

- Signature and content are coupled together
 - Easy to validate
 - Offline validation possible
- Signed/unsigned content have the same format

Enveloping Signature

- Signature contains the signed content

Consequences:

- The signature is the Parent of the content being signed
- Processing the document requires processing the signature syntax

Advantages:

- Signature and content are coupled together
 - Easy to validate
 - Offline validation possible

Disadvantages:

Signed and unsigned content have different formats → More complicated software for viewing

Detached Signature

- Signature is completely separate from the content

Consequences:

- Processing of the signature and the document are separated
- The no difference at all between signed and unsigned content

Advantages:

- If the content is a "mash up" of many disparate, separate documents, then this might be somewhat less complicated to manage

Disadvantages:

- Somehow the signature and the content must be linked together and tracked → No standards for this → Higher costs and more complex user experience
- Same disadvantage as the embedded files in PDF (discussed earlier)
- Since signed content and unsigned (non-authenticated) content are exactly the same, this could cause user confusion

Preliminary GPO Conclusions, to date:

- Current signed PDF approach uses Enveloped signatures
 - This seems to bode well for the Enveloped Signature approach
- For the case in which there is no content originator signature, this is the simplest, most effective model
 - Since there is less to go wrong and positive experience with PDF
- For cases in which a content originator signature along with disseminator signature is desired (chain of custody use cases), the Enveloped Signature approach does require a common "template" or specification between content originators and publishers
 - XML requires coordination between GPO and content originator as it is, and we already coordinate closely together on many things
- For the Chain of Custody case, it seems more complex to use either the Enveloping Signature method or the Detached Signature method
 - More complex in terms of building and maintaining software
 - Thus, to preserve potential for efficient and effective capability for the "Chain of Custody" use cases, start with Enveloped Signature method
- ENVELOPED APPROACH may be good way to start

FEEDBACK REQUESTED FROM PARTICIPANTS

Input from participants

Chain of Custody Use Cases

Chain of Custody for GPO published documents:

- Indication of the content originator → Provenance

Note that this requires:

- participation by content originators (external to GPO) GPO can't dictate
- Education process for all the content originator agencies value of adding this process at their site
 - GPO is communicating on this
- May also require coordination on common specification for signature location not unexpected, but a potential challenge

Policy and operational implications for the federal agencies

Chain of Custody Use Cases

- Factors:
 - Is ability to view the entire chain custody with the document important?
 - Open Standards based seems to be required
 - If the chain of custody is completely present/preserved with the document, is this of value?
 - For XML and PDF files, the W3C and PKCS #7 methods allow full chain of custody to travel with the document
 - Seems simpler and less complex
 - No standards exist for the protocol for the "detached " signature methods
 - The W3C and PKCS #7 methods allow full chain of custody to validated and viewed when off-line
- Other Use Cases?
- FEEDBACK REQUESTED
 - Input from Participants

Re-Authentication over time

- The Challenges:
 - #1: Standard algorithms and techniques change over long time horizons.

Examples:

- SHA-1 hash algorithm is not recommended for new signing after January 1, 2011
- 1024 bit RSA keys are not recommended for signing

#2: Issues could arise without warning on any standard algorithm or technique, requiring re-authentication using stronger/different technique.

Re-Authentication over time

Current GPO Planning:

- Maintain awareness of requirements and changes in authentication standards (continuous)
- Periodically assess requirement to re-authenticate any content (annually; more frequent as awareness of changes and requirements dictate)
- Start any re-authentication required well in advance of known requirements changes
 - Use automated, high volume authentication engines to reauthenticate
- Canvas industry and suppliers for technical systems that can make such a re-authentication process as efficient and hands-off as possible (plan to do this near term)

Granular Authentication

- The problem statement:
 - 1) Ability to authenticate an arbitrary portion of a document
 - Example: 1 page out of 500 page document
 - 2) Ability to locate text and relate it to a GPO authentic publication (published document)
- Goal:
 - An authenticated answer in a reasonable time period
- Feedback on GPO concept (next slide)
 - Concept addresses 2nd capability above (locate text)
 - Would such a concept be useful and be of value to end users?
 - Is it something GPO should investigate getting funded and built?
- FEEDBACK FROM PARTICIPANTS

Granular Authentication Concept – For Discussion



Wrap Up

Thank you for your time and participation!

GPO will post information concerning this workshop at the URL below:

http://www.gpoaccess.gov/authentication/

Information to be posted:

- today's slides (handout) (soon)
- transcript of the today's proceedings (5-10 days)
- summary report from GPO (4-6 weeks)

Wrap Up

Email comments to GPO concerning this workshop to:

authgpo@gpo.gov

Input will be accepted through July 9, 2010

The input received will be factored into the Summary Report GPO posts on this workshop